

The Bayesian ‘star paradox’ persists for long finite sequences

Mike Steel and Frederick A. Matsen

Allan Wilson Centre for Molecular Ecology and Evolution

Corresponding author:

Mike Steel

Biomathematics Research Centre

Department of Mathematics and Statistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Phone: +64-3-364-2987 ext. 7688

Fax: +64-3-364-2587

Email: M.Steel@math.canterbury.ac.nz

February 4, 2008

Keywords: phylogenetic trees, Bayesian statistics, star trees

Running head: The star paradox persists

Abstract

The ‘star paradox’ in phylogenetics is the tendency for a particular resolved tree to be sometimes strongly supported even when the data is generated by an unresolved (‘star’) tree. There have been contrary claims as to whether this phenomenon persists when very long sequences are considered. This note settles one aspect of this debate by proving mathematically that there is always a chance that a resolved tree could be strongly supported, even as the length of the sequences becomes very large.

1 Introduction

Two recent papers (Yang and Rannala 2005; Lewis, Holder and Holsinger 2005) highlighted a phenomenon that occurs when sequences evolve on a tree that contains a polytomy - in particular a three-taxon unresolved rooted tree. As longer sequences are analysed using a Bayesian approach, the posterior probability of the trees that give the different resolutions of the polytomy do not converge on relatively equal probabilities - rather a given resolution can sometimes have a posterior probability close to one. In response Kolaczkowski and Thornton (2006) investigated this phenomena further, providing some interesting simulation results, and offering an argument that seems to suggest that for very long sequences the tendency to sometimes infer strongly supported resolutions suggested by the earlier papers would disappear with sufficiently long sequences. As part of their case the authors use the expected site frequency patterns to simulate the case of infinite length sequences, concluding that “with infinite length data, posterior probabilities give equal support for all resolved trees, and the rate of false inferences falls to zero.” Of course these findings concern sequences that are effectively infinite, and, as is well known in statistics,

the limit of a function of random variables (in this case site pattern frequencies for the first n sites) does not necessarily equate with the function of the limit of the random variables. Accordingly Kolaczowski and Thornton offer this appropriate cautionary qualification of their findings:

“Analysis of ideal data sets does not indicate what will happen when very large data sets with some stochastic error are analyzed, but it does show that when infinite data are generated on a star tree, posterior probabilities are predictable, equally supporting each possible resolved tree.”

Yang and Rannala (2005) had attempted to simulate the large sample posterior distribution, but ran into numerical problems and commented that it was “unclear” what the limiting distribution on posterior probabilities was as n became large.

In particular, all of the aforementioned papers have left open an interesting statistical question, which this short note formally answers - namely, does the Bayesian posterior probability of the three resolutions of a star tree on three taxa converge to $1/3$ as the sequence length tends to infinity? That is, does the distribution on posterior probabilities for ‘very long sequences’ converge on the distribution for infinite length sequences? We show that for most reasonable priors it does not. Thus the ‘star paradox’ does not disappear as the sequences get longer.

As noted by (Yang and Rannala 2005; Lewis, Holder and Holsinger 2005) one can demonstrate such phenomena more easily for related simpler processes such as coin tossing (particularly if one imposes a particular prior). Here we avoid this simplification to avoid the criticism that such results do not rigorously establish corresponding phenomena in the phylogenetic setting, which in contrast to coin tossing involves considering a parameter space of dimension greater than 1. We also frame our main result so that it applies to a fairly general class of priors.

Note also that it is not the purpose of this short note to add to the on-going debate concerning the implications of this ‘paradox’ for Bayesian phylogenetic analysis, we merely demonstrate its existence. Some further comments and earlier references on the phenomenon have been described in the recent review paper by Alfaro and Holder 2006 (pp. 35-36).

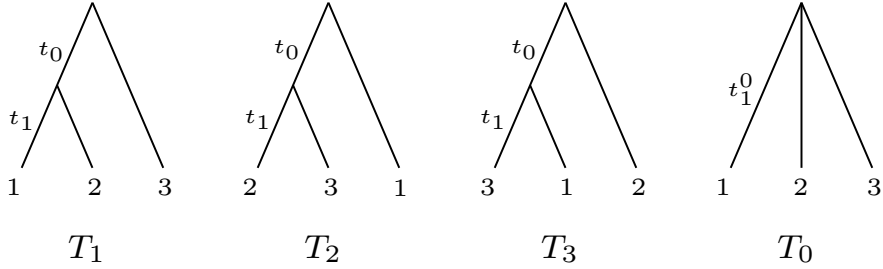


Figure 1: The three resolved rooted phylogenetic trees on three taxa T_1, T_2, T_3 , and the unresolved ‘star’ tree on which the sequences are generated T_0 .

2 Analysis of the star tree paradox for three taxa

On tree T_1 (in Fig. 1) let $p_i = p_i(t_0, t_1)$, $i = 0, 1, 2, 3$ denote the probabilities of the four site patterns (xxx, xxy, yxx, xyx , respectively) under the simple 2-state symmetric Markov process (the argument extends to more general models, but it suffices to demonstrate the phenomena for this simple model). From Eqn. (2) of (Yang and Rannala 2005) we have

$$p_0(t_0, t_1) = \frac{1}{4}(1 + e^{-4t_1} + 2e^{-4(t_0+t_1)}),$$

$$p_1(t_0, t_1) = \frac{1}{4}(1 + e^{-4t_1} - 2e^{-4(t_0+t_1)}),$$

and

$$p_2(t_0, t_1) = p_3(t_0, t_1) = \frac{1}{4}(1 - e^{-4t_1}).$$

It follows by elementary algebra that for $i = 2, 3$,

$$\frac{p_1(t_0, t_1)}{p_i(t_0, t_1)} \geq 1 + 2e^{-4t_1}(1 - e^{-4t_0}), \quad (1)$$

and thus $p_1(t_0, t_1) \geq p_i(t_0, t_1)$ with strict inequality unless $t_0 = 0$ (or in the limit as t_1 tends to infinity).

To allow maximal generality we make only minimal neutral assumptions about the prior distribution on trees and branch lengths. Namely we assume that the three resolved trees on three leaves (trees T_1, T_2, T_3 in Fig. 1) have equal prior probability $\frac{1}{3}$ and that the prior distribution on branch lengths t_0, t_1 is the same for each tree, and has a continuous joint probability density function that is everywhere non-zero. This condition applies for example to the exponential and gamma priors discussed by Yang and Rannala (2005). Any prior that satisfies these conditions we call *reasonable*. Note that we do not require that t_0 and t_1 be independent.

Let $\mathbf{n} = (n_0, n_1, n_2, n_3)$ be the counts of the different types of site patterns (corresponding to the same patterns as for the p_i 's). Thus $n = \sum_{i=0}^3 n_i$ is the total number of sites (i.e. the length of the sequences). Given a prior distribution on (t_0, t_1) for the branch lengths of T_i (for $i = 1, 2, 3$) let $\mathbb{P}[T_i|\mathbf{n}]$ be the posterior probability of tree T_i given the site pattern counts \mathbf{n} . Now suppose the n sites are generated on a star tree T_0 with positive branch lengths. We are interested in whether the posterior probability $\mathbb{P}[T_i|\mathbf{n}]$ could be close to 1 or whether the chance of generating data with this property goes to zero as the sequence length gets very large. We show that in fact the latter possibility is ruled out by our main result, namely the following:

Theorem 2.1 *Consider sequences of length n generated by a star tree T_0 on*

three taxa with strictly positive edge length t_1^0 and let \mathbf{n} be the resulting data (in terms of site pattern counts). Consider any prior on the three resolved trees (T_1, T_2, T_3) and their branch lengths that is reasonable (as defined above). For any $\epsilon > 0$, and each resolved tree T_i ($i = 1, 2, 3$), the probability that \mathbf{n} has the property that

$$\mathbb{P}(T_i | \mathbf{n}) > 1 - \epsilon$$

does not converge to 0 as n tends to infinity.

Proof of Theorem 2.1 Consider the star tree T_0 with given branch lengths t_1^0 (as in Fig. 1). Let (q_0, q_1, q_2, q_3) denote the probability of the four types of site patterns generated by T_0 with these branch lengths. Note that $q_1 = q_2 = q_3$ and so $q_0 = 1 - 3q_1$. Suppose we generate n sites on this tree, and let n_0, n_1, n_2, n_3 be the counts of the different types of site patterns (corresponding to the p_i 's). Let $\Delta_0 := \frac{n_0 - q_0 n}{\sqrt{n}}$ and for $i = 1, 2, 3$ let

$$\Delta_i := \frac{n_i - \frac{1}{3}(n - n_0)}{\sqrt{n}}.$$

For a constant $c > 1$, let F_c denote the event:

$$F_c : \Delta_2, \Delta_3 \in [-2c, -c] \text{ and } \Delta_0 \in [-c, c].$$

Notice that F_c implies $\Delta_1 \in [2c, 4c]$ since $\Delta_1 + \Delta_2 + \Delta_3 = 0$. By standard stochastic arguments (based on the asymptotic approximation of the multinomial distribution by the multinormal distribution) event F_c has probability at least some value $\delta' = \delta'(c) > 0$ for all n sufficiently large (relative to c).

Given the data $\mathbf{n} = (n_0, n_1, n_2, n_3)$ the assumption of equality of priors across T_1, T_2 and T_3 implies that

$$\mathbb{P}(n_0, n_1, n_2, n_3 | T_2, t_0, t_1) = \mathbb{P}(n_0, n_2, n_3, n_1 | T_1, t_0, t_1), \quad (2)$$

and

$$\mathbb{P}(n_0, n_1, n_2, n_3 | T_3, t_0, t_1) = \mathbb{P}(n_0, n_3, n_1, n_2 | T_1, t_0, t_1). \quad (3)$$

Now, as (t_0, t_1) are random variables with some prior density, when we view p_0, p_1, p_2, p_3 as random variables by virtue of their dependence on (t_0, t_1) , we will write them as P_0, P_1, P_2, P_3 (note that Yang and Rannala (2005) use P_i differently). With this notation, the posterior probability of T_1 conditional on \mathbf{n} can be written as

$$\mathbb{P}(T_1|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_1} P_2^{n_2} P_3^{n_3}]$$

where $p(\mathbf{n})$ is the posterior probability of \mathbf{n} and \mathbb{E}_1 denotes expectation with respect to the prior for t_0, t_1 on T_1 . Moreover since $P_2 = P_3$, we can write this as $\mathbb{P}(T_1|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_1} P_2^{n_2+n_3}]$. By (2) and (3) we have

$$\mathbb{P}(T_2|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_1} P_2^{n_1+n_3}]; \text{ and } \mathbb{P}(T_3|\mathbf{n}) = p(\mathbf{n})^{-1} \times \mathbb{E}_1[P_0^{n_0} P_1^{n_1} P_2^{n_1+n_2}]$$

where again expectation is taken with respect to the prior for t_0, t_1 on T_1 . Consequently,

$$\frac{\mathbb{P}(T_1|\mathbf{n})}{\mathbb{P}(T_2|\mathbf{n})} = \frac{\mathbb{E}_1[X]}{\mathbb{E}_1[Y]}, \quad (4)$$

where

$$X = P_0^{n_0} P_1^{n_1} P_2^{n_2+n_3} \text{ and } Y = P_0^{n_0} P_1^{n_1} P_2^{n_1+n_3}.$$

As will be shown later, it suffices to demonstrate that the ratio in (4) can be large with nonvanishing probability in order to obtain the conclusion of the theorem. In order to do so we use the following lemma, whose proof is provided in the Appendix.

Lemma 2.2 *Let X, Y be non-negative continuous random variables, dependent on a third random variable Λ that takes values in an interval $I = [a, b]$. Suppose that for some interval I_0 strictly inside I , and $I_1 = I - I_0$ the following inequality holds:*

$$\mathbb{E}[Y|\Lambda \in I_0] \geq \mathbb{E}[Y|\Lambda \in I_1], \quad (5)$$

and that for some constant $B > 0$, and all $\lambda \in I_0$,

$$\frac{\mathbb{E}[X|\Lambda = \lambda]}{\mathbb{E}[Y|\Lambda = \lambda]} \geq B. \quad (6)$$

Then, $\frac{\mathbb{E}[X]}{\mathbb{E}[Y]} \geq B \cdot \mathbb{P}(\Lambda \in I_0)$.

To apply this lemma, select a value $s > 0$ so that $\frac{1}{4} + s < q_0 < 1 - s$, and let $I_0 = [q_0 - s, q_0 + s]$. Then let $I = [\frac{1}{4}, 1]$ and $I_1 = I - I_0$.

Claim: For n sufficiently large, and conditional on the data $\mathbf{n} = (n_0, n_1, n_2, n_3)$ satisfying F_c :

$$(i) \quad \mathbb{E}_1[Y|P_0 \in I_0] \geq \mathbb{E}_1[Y|P_0 \in I_1]$$

$$(ii) \quad \text{For all } p_0 \in I_0, \quad \frac{\mathbb{E}_1[X|P_0=p_0]}{\mathbb{E}_1[Y|P_0=p_0]} \geq 6c^2.$$

The proofs of these two claims is given in the Appendix.

Applying Lemma 2.2 to the Claims (i) and (ii) we deduce that conditional on \mathbf{n} satisfying F_c and n being sufficiently large,

$$\frac{\mathbb{E}_1[X]}{\mathbb{E}_1[Y]} \geq 6c^2 \cdot \mathbb{P}(P_0 \in I_0). \quad (7)$$

Select $c > \frac{1}{\sqrt{3\epsilon\mathbb{P}(P_0 \in I_0)}}$ (this is finite by the assumption that the prior on (t_0, t_1) is everywhere non-zero). As stated before, the probability that \mathbf{n} satisfies F_c is at least $\delta' = \delta'(c) > 0$ for n sufficiently large. Then, $6c^2 \cdot \mathbb{P}(P_0 \in I_0) > \frac{2}{\epsilon}$ and so by (7), $\frac{\mathbb{P}(T_1|\mathbf{n})}{\mathbb{P}(T_2|\mathbf{n})} = \frac{\mathbb{E}_1[X]}{\mathbb{E}_1[Y]} > \frac{2}{\epsilon}$. Similarly, $\frac{\mathbb{P}(T_1|\mathbf{n})}{\mathbb{P}(T_3|\mathbf{n})} > \frac{2}{\epsilon}$. Now, since $\mathbb{P}(T_1|\mathbf{n}) + \mathbb{P}(T_2|\mathbf{n}) + \mathbb{P}(T_3|\mathbf{n}) = 1$ it now follows that, for n sufficiently large, and conditional on an event of probability at least $\delta' > 0$, that $\mathbb{P}(T_1|\mathbf{n}) > 1 - \epsilon$ as claimed. This completes the proof. \square

2.1 Concluding remarks

One feature of the argument we have provided is that it does not require stipulating in advance a particular prior on the branch lengths – that is, the result is somewhat generic as it imposes relatively few conditions. Moreover, the requirement that the prior on (t_0, t_1) be everywhere non-zero could be weakened to simply being non-zero in a neighborhood of $(0, t_1^0)$ (thereby allowing, for example, a uniform distribution on bounded range).

A interesting open question in the spirit of this paper is to explicitly calculate the limit of the posterior density $f(P_1, P_2, P_3)$ described in (Yang and Rannala 2005).

2.2 Acknowledgments

MS thanks Ziheng Yang for suggesting the problem of computing the limiting distribution of posterior probabilities for 3-taxon trees. This work is funded by the *Allan Wilson Centre for Molecular Ecology and Evolution*.

References

- Alfaro ME, Holder MT. 2006. The posterior and prior in Bayesian phylogenetics. *Annu. Rev. Evol. Syst.* 37: 19-42.
- Kolaczowski B, Thornton JW. 2006. Is there a star tree paradox? *Mol. Biol. Evol.* 23: 1819–1823.
- Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. *Syst. Biol.* 54 (2): 241-253.

Yang Z, Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.* 54 (3): 455-470.

3 Appendix: Proof of Lemma 2.2 and Claims (i), (ii)

Proof of Lemma 2.2: For $W = X, Y$ we have

$$\mathbb{E}[W] = \mathbb{E}[W|\Lambda \in I_0]\mathbb{P}(\Lambda \in I_0) + \mathbb{E}[W|\Lambda \in I_1]\mathbb{P}(\Lambda \in I_1). \quad (8)$$

In particular, for $W = X$ we have: $\mathbb{E}[X] \geq \mathbb{E}[X|\Lambda \in I_0]\mathbb{P}(\Lambda \in I_0)$. Note that (6) implies that $\mathbb{E}[X|\Lambda \in I_0] \geq B \cdot \mathbb{E}[Y|\Lambda \in I_0]$, so

$$\mathbb{E}[X] \geq B \cdot \mathbb{E}[Y|\Lambda \in I_0]\mathbb{P}(\Lambda \in I_0). \quad (9)$$

Taking $W = Y$ in (8) and applying (5) gives us

$$\mathbb{E}[Y] \leq \mathbb{E}[Y|\Lambda \in I_0](\mathbb{P}(\Lambda \in I_0) + \mathbb{P}(\Lambda \in I_1)) = \mathbb{E}[Y|\Lambda \in I_0]$$

which combined with (9) gives the result. \square

Proof of Claim (i), $\mathbb{E}_1[Y|P_0 \in I_0] \geq \mathbb{E}_1[Y|P_0 \in I_1]$:

We will first bound $\mathbb{E}_1[Y|P_0 \in I_1]$ above. Let $\mu(n) = (q_0^{q_0} q_1^{q_1} q_2^{q_2} q_3^{q_3})^n$. Now, conditional on \mathbf{n} satisfying F_c we have

$$n^{-1} \log(\mu(n)/Y(t_0, t_1)) = d_{KL}(q, p) + o(1),$$

where $p = (p_0, p_1, p_2, p_3)$ and $q = (q_0, q_1, q_2, q_3)$, and d_{KL} denotes Kullback-Leibler distance. Now, $d_{KL}(q, p) \geq \frac{1}{2}\|q - p\|_1^2 \geq \frac{1}{2}|q_0 - p_0|^2$ (the first inequality is a standard one in probability theory). In particular, if $p_0 \in I_1$, then $|q_0 - p_0| > s > 0$. Moreover,

$$\mathbb{E}_1[Y|P_0 \in I_1] \leq \max\{Y(t_0, t_1) : p_0(t_0, t_1) \in I_1\}.$$

Summarizing,

$$\mathbb{E}_1[Y|P_0 \in I_1] \leq \max\{Y(t_0, t_1) : p_0(t_0, t_1) \in I_1\} < \mu(n)e^{-\frac{1}{2}s^2n+o(n)}. \quad (10)$$

In the reverse direction, we have:

$$\mathbb{E}_1[Y|P_0 \in I_0] \geq A(n)B(n)$$

where

$$A(n) = \min \{Y(t_0, t_1) : (t_0, t_1) \in [0, n^{-1}] \times [t_1^0, t_1^0 + n^{-1}]\}$$

and

$$B(n) = \mathbb{P}((t_0, t_1) \in [0, n^{-1}] \times [t_1^0, t_1^0 + n^{-1}]).$$

Now,

$$A(n)/\mu(n) = \left(\frac{p_0^{q_0} p_1^{q_1} p_2^{2q_1}}{q_0^{q_0} q_1^{3q_1}} \right)^n \cdot (p_1^{\Delta_2 - \frac{1}{3}\Delta_0} p_2^{\Delta_1 + \Delta_3 - \frac{2}{3}\Delta_0})^{\sqrt{n}}.$$

Now, the first term of this product converges to a constant as n grows (because $p_0 - q_0, p_1 - q_1$ and $p_2 - q_1$ are each of order n^{-1}) while the condition F_c ensures that the second term decays no faster than $e^{-C_1\sqrt{n}}$ for a constant C_1 . Thus, $A(n) \geq C_2\mu(n)e^{-C_1\sqrt{n}}$ for a positive constant C_2 . The term $B(n)$ is asymptotically proportional to n^{-2} . Summarizing, for a constant $C_3 > 0$ (dependent just on t_1^0)

$$\mathbb{E}_1[Y|P_0 \in I_0] \geq C_3\mu(n)n^{-2}e^{-C_1\sqrt{n}},$$

which combined with (10) establishes claim (i) for n sufficiently large. \square

In order to prove claim (ii) we need some preliminary results.

Lemma 3.1 *Let $\eta < 1$. Then for each $x > 0$ there exists a value $K = K(x) < \infty$ that depends continuously on x so that the following holds. For any continuous random variable Z on $[0, 1]$ with a smooth density function f that satisfies $f(1) \neq 0$ and $|f'(z)| < B$ for all $z \in (\eta, 1]$, we have*

$$k \cdot \frac{(\mathbb{E}[Z^k] - \mathbb{E}[Z^{k+1}])}{\mathbb{E}[Z^k]} \geq \frac{1}{2}$$

for all $k \geq K(\frac{B}{f(1)})$.

Proof. Let $t_k = 1 - \frac{1}{\sqrt{k}}$. Then

$$\mathbb{E}[Z^k] = \int_0^{t_k} t^k f(t) dt + \int_{t_k}^1 t^k f(t) dt.$$

Now,

$$0 \leq \int_0^{t_k} t^k f(t) dt \leq t_k^k \sim e^{-\sqrt{k}-1/2},$$

where \sim denotes asymptotic equivalence (i.e. $f(k) \sim g(k)$ iff $\lim_{k \rightarrow \infty} f(k)/g(k) = 1$). Using integration by parts,

$$\int_{t_k}^1 t^k f(t) dt = \frac{1}{k+1} t^{k+1} f(t) \Big|_{t_k}^1 - \frac{1}{k+1} \int_{t_k}^1 t^{k+1} f'(t) dt.$$

Now, provided $k > (1 - \eta)^{-2}$ we have $t_k > \eta$ and so the absolute value of the second term on the right is at most $\frac{B}{k+1} \int_{t_k}^1 t^{k+1} dt = \frac{B}{(k+1)(k+2)} (1 - t_k^{k+2})$. Consequently, $\left| \mathbb{E}[Z^k] - \frac{f(1)}{k+1} \right|$ is bounded above by B times a term of order k^{-2} . A similar argument, again using integration by parts, shows that $\left| k(\mathbb{E}[Z^k] - \mathbb{E}[Z^{k+1}]) - \frac{f(1)}{k+1} \right|$ is bounded above by B times a term of order k^{-2} , and the lemma now follows by some routine analysis. \square

Lemma 3.2 *Let $y = (1 + 2x)(1 - x)^2$. Then for $x \in [0, 1)$ and $m \geq 3$ we have*

$$\left(\frac{1 + 2x}{1 - x} \right)^m \geq m^2(1 - y).$$

Proof.

$$\left(\frac{1 + 2x}{1 - x} \right)^m = \left(1 + \frac{3x}{1 - x} \right)^m \geq \frac{m(m-1)}{2} \left(\frac{3x}{1 - x} \right)^2 \geq \frac{9m(m-1)x^2}{2},$$

and $m^2(1 - y) = m^2(3x^2 - 2x^3) \leq 3m^2x^2$ and for $m \geq 3$ this upper bound is less than the lower bound in the previous expression. \square

Proof of Claim (ii), for all $p_0 \in I_0$, $\frac{\mathbb{E}_1[X|P_0=p_0]}{\mathbb{E}_1[Y|P_0=p_0]} \geq 6c^2$:

Write $\mathbb{E}_1[W|p_0]$ as shorthand for $\mathbb{E}[W|P_0 = p_0]$. Note that, for any $r, s > 0$, $\mathbb{E}_1[P_0^{n_0} P_1^r P_2^s | p_0] = p_0^{n_0} \mathbb{E}_1[P_1^r P_2^s | p_0]$. Consequently, if we let $k = k(n) = \frac{1}{3}(n -$

n_0) then, by definition of the Δ_i 's,

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} = \frac{\mathbb{E}_1[(P_1 P_2^2)^k \cdot (P_1^{\Delta_1} P_2^{\Delta_2 + \Delta_3})^{\sqrt{n}} | p_0]}{\mathbb{E}_1[(P_1 P_2^2)^k \cdot (P_1^{\Delta_2} P_2^{\Delta_1 + \Delta_3})^{\sqrt{n}} | p_0]}. \quad (11)$$

Now, conditional on \mathbf{n} satisfying F_c (and since $P_1 \geq P_2$) the following two inequalities hold

$$P_1^{\Delta_1} P_2^{\Delta_2 + \Delta_3} = \left(\frac{P_1}{P_2}\right)^{\Delta_1} \geq \left(\frac{P_1}{P_2}\right)^{2c} \quad \text{and} \quad P_1^{\Delta_2} P_2^{\Delta_1 + \Delta_3} = \left(\frac{P_1}{P_2}\right)^{\Delta_2} \leq 1.$$

Applying this to (11) gives:

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq \frac{\mathbb{E}_1 \left[(P_1 P_2^2)^k \cdot \left(\frac{P_1}{P_2}\right)^{2c\sqrt{n}} | p_0 \right]}{\mathbb{E}_1[(P_1 P_2^2)^k | p_0]}. \quad (12)$$

Let $U = \frac{P_1 - P_2}{1 - P_0}$, which takes values between 0 and 1 because $P_1 \geq P_2$. Since $P_1 + 2P_2 = 1 - P_0$, we can write $P_1 = \frac{1}{3}(1 + 2U)(1 - P_0)$ and $P_2 = \frac{1}{3}(1 - U)(1 - P_0)$. Thus, $P_1 P_2^2 = \frac{1}{27}(1 + 2U)(1 - U)^2(1 - P_0)^3$ and $\frac{P_1}{P_2} = \frac{(1 + 2U)}{(1 - U)}$. Substituting these into (12), letting $Z = (1 + 2U)(1 - U)^2$ and noting that $\sqrt{n} \geq \sqrt{3k}$ gives

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq \frac{\mathbb{E}_1 \left[Z^k \cdot \left(\frac{1 + 2U}{1 - U}\right)^{2c\sqrt{3k}} | p_0 \right]}{\mathbb{E}_1[Z^k | p_0]}.$$

Thus, by Lemma 3.2, (taking $x = U, y = Z, m = 2c\sqrt{3k}$) we obtain, for $m \geq 3$,

$$\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq 12c^2 k \cdot \frac{(\mathbb{E}_1[Z^k | p_0] - \mathbb{E}_1[Z^{k+1} | p_0])}{\mathbb{E}_1[Z^k | p_0]}. \quad (13)$$

Now the mapping $(t_0, t_1) \mapsto (P_0, Z)$ is a smooth invertible mapping between $(0, \infty)^2$ and its image within $(\frac{1}{4}, 1) \times (0, 1)$. Notice that Z approaches 1 whenever P_0 approaches $\frac{1}{4}$ or 1 (in particular, even if t_0, t_1 are independent, P_0 and Z generally will not be). However over the interval I_0 the conditional density $f(Z|P_0 = p_0)$ of Z given a value p_0 for P_0 is smooth and bounded away from 0, and its first derivative is also bounded above over this interval. Consequently, we may apply Lemma 3.1 (noting that the condition that \mathbf{n} satisfies F_c ensures that $k(n) \geq \frac{1}{4}n - o(n)$) to show that for n sufficiently large the following inequality holds for all $p_0 \in I_0$,

$$k \cdot \frac{(\mathbb{E}_1[Z^k | p_0] - \mathbb{E}_1[Z^{k+1} | p_0])}{\mathbb{E}_1[Z^k | p_0]} \geq \frac{1}{2}.$$

Applying this to (13) gives $\frac{\mathbb{E}_1[X|p_0]}{\mathbb{E}_1[Y|p_0]} \geq 6c^2$ as claimed. This completes the proof of Claim (ii).